Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

Evaluasi Validitas Model Machine learning pada Klasifikasi Stunting Berbasis Data Antropometri dan Hubungan Deterministik

Turwan Aldi Putra*, Nirwana Hendrastuty

Fakultas Teknik dan Ilmu Komputer, Prodi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1,*]turwan_aldi_putra@teknokrat.ac.id, [2]nirwanahendrastuty@teknokrat.ac.id
Email Penulis Korespondensi: turwan_aldi_putra@teknokrat.ac.id
Submitted: **03/04/2026**; Accepted: **02/06/2026**; Published**: 05/06/2026**

Abstrak−Stunting merupakan permasalahan gizi kronis pada balita yang berdampak pada pertumbuhan dan perkembangan anak. Berbagai penelitian telah memanfaatkan machine learning untuk klasifikasi status gizi berbasis data antropometri, namun validitas model yang dihasilkan masih jarang dikaji. Penelitian ini bertujuan untuk mengevaluasi validitas model machine learning dalam klasifikasi status stunting menggunakan algoritma XGBoost, Random Forest, dan Naïve Bayes. Dataset yang digunakan merupakan data antropometri balita sebanyak 120.999 data dengan atribut umur, jenis kelamin, dan tinggi badan sebagai fitur, serta status gizi sebagai variabel target. Proses penelitian meliputi tahap preprocessing, transformasi data, serta evaluasi model menggunakan metode k-fold cross-validation dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa Random Forest dan XGBoost menghasilkan akurasi yang sangat tinggi, masing-masing sebesar 99,91% dan 99,08%, sedangkan Naïve Bayes hanya mencapai 55%. Perbedaan performa yang kontras ini menunjukkan bahwa model berbasis ensemble mampu menangkap pola yang sangat kuat dalam data, sementara Naïve Bayes mengalami kesulitan akibat keterkaitan antar fitur. Selain itu, tingginya akurasi pada model tertentu mengindikasikan adanya hubungan deterministik antara fitur dan label, yang berpotensi menyebabkan model kurang robust terhadap data yang mengandung kesalahan pengukuran atau noise.

Kata Kunci: Data Antropometri; Klasifikasi; Machine Learning; Stunting; Validitas Model

Abstract−Stunting is a chronic nutritional problem among infants and toddlers that affects children's growth and development. Various studies have utilized machine learning for nutritional status classification based on anthropometric data; however, the validity of the resulting models has rarely been examined. This study aims to evaluate the validity of machine learning models in classifying stunting status using the XGBoost, Random Forest, and Naïve Bayes algorithms. The dataset consists of 120,999 anthropometric records of infants, with age, gender, and height as features, and nutritional status as the target variable. The research process included preprocessing, data transformation, and model evaluation using the k-fold cross-validation method with accuracy, precision, recall, and F1-score metrics. The results showed that Random Forest and XGBoost achieved very high accuracy, at 99.91% and 99.08%, respectively, while Naïve Bayes reached only 55%. This stark difference in performance indicates that ensemble-based models are capable of capturing very strong patterns in the data, while Naïve Bayes struggles due to the interdependence among features. Furthermore, the high accuracy of certain models suggests a deterministic relationship between features and labels, which could potentially make the models less robust against data containing measurement errors or noise.

Keywords: Anthropometric Data; Classification; Machine Learning; Stunting; Model Validity

1. PENDAHULUAN

Stunting merupakan salah satu permasalahan kesehatan yang masih menjadi perhatian utama di berbagai negara berkembang, termasuk Indonesia. Kondisi ini tidak hanya berdampak pada pertumbuhan fisik anak, tetapi juga pada perkembangan kognitif dan produktivitas jangka panjang [1] . *World Health Organization* (WHO) menyatakan bahwa penilaian status stunting dilakukan berdasarkan indikator antropometri, khususnya hubungan antara tinggi badan dan umur yang mengacu pada standar pertumbuhan anak [2]. Oleh karena itu, data antropometri menjadi dasar utama dalam menentukan status gizi balita.

Dalam beberapa tahun terakhir, pendekatan *machine learning* mulai banyak digunakan untuk membantu klasifikasi status stunting. Berbagai algoritma seperti Random Forest, Support Vector Machine, dan Extreme Gradient Boosting dilaporkan mampu menghasilkan akurasi yang tinggi dalam memprediksi status gizi anak [3],[4],[5]. Selain itu, metode berbasis *ensemble learning* juga diklaim dapat meningkatkan performa model secara signifikan [6]. Hasil-hasil tersebut menunjukkan bahwa *machine learning* memiliki potensi besar dalam mendukung analisis data kesehatan, termasuk dalam deteksi stunting [7].

Namun demikian, sebagian besar penelitian sebelumnya masih berfokus pada perbandingan performa algoritma, terutama dalam hal akurasi, tanpa mengevaluasi apakah model yang dihasilkan benar-benar valid secara konseptual. Pendekatan yang hanya berorientasi pada akurasi berpotensi menimbulkan *bias* dalam interpretasi hasil, karena model dengan performa tinggi belum tentu memiliki kemampuan generalisasi yang baik [8]. Dalam domain kesehatan, hal ini menjadi masalah serius karena keputusan yang diambil berdasarkan model yang tidak valid dapat berdampak langsung pada penanganan pasien.

Beberapa penelitian menunjukkan bahwa model *machine learning* rentan terhadap berbagai permasalahan seperti *data leakage*, *bias*, dan perbedaan distribusi data antara pelatihan dan implementasi nyata [9], [10], [11]. *Data leakage*, misalnya, dapat terjadi ketika informasi yang digunakan untuk menentukan label secara tidak langsung juga digunakan sebagai fitur, sehingga model hanya mempelajari hubungan yang bersifat langsung dan tidak realistis [12]. Kondisi ini menyebabkan performa model terlihat sangat tinggi, tetapi sebenarnya tidak mencerminkan kemampuan prediksi yang sesungguhnya. Selain itu, pedoman TRIPOD+AI menekankan bahwa evaluasi model dalam bidang

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

kesehatan harus mempertimbangkan validitas, transparansi, dan kemampuan generalisasi, bukan hanya metrik performa semata [13].

Permasalahan ini menjadi semakin relevan dalam klasifikasi stunting berbasis data antropometri. [14] Status stunting ditentukan berdasarkan perhitungan tinggi badan terhadap umur, sehingga variabel yang digunakan sebagai fitur memiliki hubungan langsung dengan label. Hal ini berpotensi menimbulkan hubungan deterministik, di mana model tidak benar-benar belajar dari data, tetapi hanya merekonstruksi aturan yang telah ada. Dalam kondisi seperti ini, algoritma dengan kemampuan kompleks seperti Random Forest atau XGBoost cenderung menghasilkan akurasi yang sangat tinggi, sementara algoritma sederhana seperti Naïve Bayes menunjukkan performa yang lebih rendah. Perbedaan ini perlu dianalisis secara kritis untuk memahami apakah model benar-benar melakukan proses pembelajaran atau hanya memanfaatkan struktur data yang bersifat deterministik.

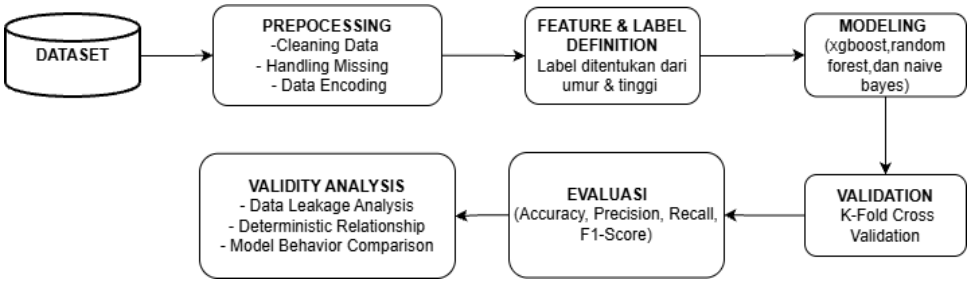
Selain itu, penggunaan *dataset* sekunder seperti Kaggle dalam penelitian *machine learning* juga perlu mendapatkan perhatian khusus [15]. *Dataset* publik umumnya telah melalui proses pembersihan data, sehingga memiliki tingkat noise yang lebih rendah dibandingkan dengan data kesehatan di lapangan. Dalam praktik nyata, data antropometri sering mengandung kesalahan pengukuran, missing values, dan variasi yang tinggi akibat faktor manusia dan lingkungan. Perbedaan ini dapat menyebabkan model yang dilatih pada *dataset* yang bersih tidak mampu melakukan generalisasi dengan baik ketika diterapkan pada kondisi nyata [11]. Oleh karena itu, penting untuk mengevaluasi validitas model dengan mempertimbangkan karakteristik *dataset* yang digunakan.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengevaluasi validitas model *machine learning* dalam klasifikasi status stunting berbasis data antropometri. Evaluasi dilakukan dengan mengkaji potensi *data leakage*, hubungan deterministik antara fitur dan label, serta implikasinya terhadap performa model. Studi ini menggunakan algoritma XGBoost, Random Forest, dan Naïve Bayes sebagai perbandingan untuk memahami bagaimana perbedaan karakteristik algoritma mempengaruhi hasil klasifikasi. Dengan demikian, penelitian ini diharapkan dapat memberikan pemahaman yang lebih kritis mengenai penggunaan *machine learning* dalam domain kesehatan, khususnya dalam klasifikasi status stunting.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan sistematis untuk mengevaluasi validitas model *machine learning* dalam klasifikasi status stunting berbasis data antropometri. Berbeda dengan penelitian konvensional yang berfokus pada pencapaian performa model, penelitian ini menitikberatkan pada analisis kritis terhadap hubungan antara fitur dan label serta implikasinya terhadap hasil klasifikasi. Alur lengkap tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Diagram Flow

Proses tahapan penelitian dimulai dari proses pengumpulan *dataset* yang digunakan sebagai sumber data utama. *Dataset* yang digunakan diperoleh dari platform Kaggle dan memuat atribut antropometri balita, seperti umur, tinggi badan, jenis kelamin, serta status gizi yang digunakan sebagai label [16]. Data yang diperoleh kemudian melalui tahapan preprocessing untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam proses pemodelan. Tahapan ini meliputi data cleaning untuk menghilangkan data yang tidak valid, penanganan missing value, serta proses encoding untuk mengubah data kategorikal menjadi representasi numerik. Selanjutnya dilakukan tahap penentuan fitur dan label. Fitur yang digunakan dalam penelitian ini meliputi umur, tinggi badan, dan jenis kelamin, sedangkan label berupa status gizi yang ditentukan berdasarkan standar World Health Organization (WHO). Pada tahap ini diperhatikan bahwa label status gizi diturunkan langsung dari variabel umur dan tinggi badan, sehingga terdapat potensi hubungan deterministik antara fitur dan label yang dapat mempengaruhi hasil klasifikasi. Tahap berikutnya adalah proses pemodelan menggunakan algoritma *machine learning*, yaitu XGBoost, Random Forest, dan Naïve Bayes [17]. Untuk memastikan hasil evaluasi yang lebih reliabel dan menghindari *bias* dari pembagian data tunggal, penelitian ini tidak menggunakan metode *train-test split* konvensional, melainkan menerapkan metode *k-fold cross-validation* dalam proses evaluasi model [13]. Evaluasi model dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* [5]. Namun, dalam penelitian ini metrik performa tidak digunakan sebagai satu-satunya indikator, melainkan sebagai dasar untuk analisis lebih lanjut. Tahap akhir penelitian adalah analisis

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

validitas model yang meliputi identifikasi potensi *data leakage*, analisis hubungan deterministik antara fitur dan label, serta perbandingan perilaku model dalam mempelajari pola data. Analisis ini bertujuan untuk menilai apakah performa model mencerminkan kemampuan prediksi yang sebenarnya atau hanya dipengaruhi oleh struktur data yang digunakan. Seluruh tahapan penelitian divisualisasikan pada Gambar 1.

2.2 *Dataset*

Dataset yang digunakan dalam penelitian ini merupakan *dataset* status gizi balita yang diperoleh dari platform Kaggle, yang menyediakan berbagai *dataset* publik untuk keperluan analisis dan penelitian. *Dataset* dapat diakses melalui https://www.kaggle.com/*datasets*/rendiputra/stunting-balita-detection-121k-rows. *Dataset* ini termasuk dalam kategori data sekunder yang bersifat terbuka dan umum digunakan dalam penelitian berbasis *machine learning*. *Dataset* terdiri dari 120.999 baris data dengan empat atribut utama, yaitu Umur (bulan), Jenis Kelamin, Tinggi Badan (cm), dan Status Gizi. Variabel Umur (bulan) merepresentasikan usia balita dalam satuan bulan, sedangkan Tinggi Badan (cm) merupakan indikator utama dalam menilai pertumbuhan anak. Variabel Jenis Kelamin digunakan sebagai atribut tambahan yang dapat mempengaruhi pola pertumbuhan. Adapun variabel Status Gizi merupakan variabel target yang terdiri dari beberapa kategori, yaitu normal, stunted, severely stunted, dan tinggi, sehingga penelitian ini termasuk dalam klasifikasi multi-kelas [5]. Penentuan label Status Gizi dalam *dataset* ini didasarkan pada standar antropometri *World Health Organization* (WHO), khususnya indikator *Height-for-Age Z-score* (HAZ) [2]. Dengan demikian, label yang digunakan dalam penelitian ini secara langsung diturunkan dari variabel Umur dan Tinggi Badan. Kondisi ini menunjukkan adanya potensi hubungan deterministik antara fitur dan label, yang dapat menyebabkan model *machine learning* tidak melakukan proses pembelajaran secara independen, melainkan hanya merekonstruksi aturan klasifikasi yang telah ditentukan [12]. Selain itu, sebagai *dataset* publik, data yang digunakan kemungkinan telah melalui proses kurasi dan pembersihan, sehingga memiliki tingkat noise yang relatif rendah dibandingkan dengan data kesehatan di lapangan. Dalam praktik nyata, [15] data antropometri seringkali mengandung kesalahan pengukuran, missing values, serta variasi yang tinggi akibat faktor manusia dan lingkungan. Perbedaan karakteristik ini berpotensi mempengaruhi kemampuan generalisasi model ketika diterapkan pada kondisi dunia nyata [11]. Oleh karena itu, dalam penelitian ini *dataset* tidak hanya digunakan sebagai sumber data untuk pelatihan model, tetapi juga sebagai objek kajian untuk mengevaluasi validitas pendekatan *machine learning* dalam klasifikasi status stunting berbasis data antropometri.

2.3 *Preprocessing*

Tahap preprocessing merupakan proses penting dalam penelitian ini untuk memastikan bahwa *dataset* berada dalam kondisi yang layak digunakan dalam proses pemodelan *machine learning*. Berbeda dengan pendekatan konvensional yang berfokus pada peningkatan performa model, tahap preprocessing dalam penelitian ini dilakukan secara minimal untuk mempertahankan struktur asli data, sehingga memungkinkan analisis terhadap validitas model dan hubungan antara fitur dan label [18] .

Tahapan preprocessing yang dilakukan meliputi beberapa proses utama, yaitu data cleaning, penanganan missing value, dan encoding data.

a. *Data cleaning*
   Langkah pertama data cleaning dilakukan untuk memastikan bahwa *dataset* tidak mengandung data yang tidak valid atau inkonsisten. Proses ini meliputi penghapusan data duplikat, pengecekan kesesuaian tipe data, serta eliminasi nilai yang tidak logis, seperti nilai tinggi badan nol atau negatif. Tahapan ini penting untuk menghindari kesalahan dalam proses pelatihan model dan menjaga kualitas data [9].
b. *Handling Missing value*
   Selanjutnya dilakukan proses identifikasi dan penanganan missing value pada *dataset*. Apabila ditemukan data yang tidak lengkap, maka dilakukan penanganan dengan metode yang sesuai, seperti penghapusan data atau imputasi sederhana. Penanganan missing value bertujuan untuk menjaga konsistensi data tanpa mengubah karakteristik utama *dataset*, sehingga tidak mengganggu interpretasi model [19].
c. *Encoding*
   data *encoding* untuk mengubah data kategorikal Tahap encoding dilakukan untuk mengubah data kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Variabel jenis kelamin yang merupakan data kategorikal diubah menjadi representasi numerik menggunakan metode encoding sederhana.
   Sementara itu, variabel status gizi sebagai label target juga direpresentasikan dalam bentuk numerik untuk keperluan pemodelan. Namun, perlu diperhatikan bahwa label ini telah ditentukan sebelumnya berdasarkan standar WHO, sehingga proses encoding tidak mengubah makna atau struktur label.
d. *Methodological Considerations for Model Validity*
   Dalam penelitian ini tidak dilakukan proses *feature selection* maupun teknik peningkatan data seperti *oversampling*. Hal ini dikarenakan penggunaan teknik tersebut dapat mengubah distribusi data dan berpotensi mempengaruhi validitas hasil, terutama dalam konteks data yang memiliki hubungan langsung antara fitur dan label [20]. Selain itu, penelitian ini juga tidak berfokus pada optimasi model melalui tuning parameter secara ekstensif, melainkan pada analisis bagaimana model merespon struktur data yang ada. Pendekatan ini sejalan dengan pandangan bahwa model *machine learning* tidak hanya perlu dinilai dari performa prediksi, tetapi juga dari kemampuan interpretasi dan validitasnya dalam merepresentasikan fenomena yang dikaji [21].

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

2.4 *Feature and Label Definition*

Pada penelitian ini, proses penentuan fitur dan label dilakukan berdasarkan karakteristik data antropometri yang digunakan dalam klasifikasi status stunting. Fitur yang digunakan dalam penelitian ini meliputi tiga variabel utama, yaitu Umur (bulan), Tinggi Badan (cm), dan Jenis Kelamin. Variabel-variabel tersebut dipilih karena secara umum digunakan dalam analisis pertumbuhan anak dan sering menjadi dasar dalam penelitian terkait klasifikasi status gizi balita [22]. Sementara itu, variabel target yang digunakan adalah Status Gizi, yang terdiri dari beberapa kategori yaitu normal, stunted, severely stunted, dan tinggi. Penentuan label ini didasarkan pada standar antropometri World Health Organization (WHO), khususnya indikator Height-for-Age Z-score (HAZ) [2]. Namun demikian, perlu diperhatikan bahwa label Status Gizi dalam *dataset* ini secara langsung diturunkan dari variabel Umur dan Tinggi Badan. Hal ini menunjukkan adanya hubungan deterministik antara fitur dan label, di mana nilai label sepenuhnya ditentukan oleh kombinasi nilai fitur tersebut.

Dalam kondisi ini, model *machine learning* tidak sepenuhnya melakukan proses pembelajaran terhadap pola yang independen, melainkan berpotensi hanya merekonstruksi aturan klasifikasi yang telah ditentukan sebelumnya. Kondisi tersebut memiliki implikasi penting terhadap validitas model yang dihasilkan. Model dengan performa tinggi dalam konteks ini belum tentu mencerminkan kemampuan prediksi yang sebenarnya, melainkan dapat disebabkan oleh kemampuan model dalam menangkap hubungan langsung antara fitur dan label. Fenomena ini berkaitan dengan konsep *data leakage*, di mana informasi yang digunakan untuk membentuk label secara tidak langsung tersedia dalam fitur yang digunakan untuk pelatihan model [18]. Selain itu, penggunaan fitur yang secara langsung berkaitan dengan pembentukan label juga dapat mempengaruhi interpretasi hasil model, terutama dalam konteks penerapan di dunia nyata. Model yang dibangun dari data dengan hubungan deterministik cenderung memiliki keterbatasan dalam melakukan generalisasi pada data baru yang memiliki karakteristik berbeda [18]. Oleh karena itu, dalam penelitian ini penentuan fitur dan label tidak hanya dipandang sebagai bagian dari proses teknis, tetapi juga sebagai aspek penting dalam evaluasi validitas model *machine learning* pada klasifikasi status stunting berbasis data antropometri.

2.5 Modeling dan Evaluation

Pada penelitian ini, proses pengembangan model dilakukan dengan menggunakan beberapa algoritma *machine learning* yang memiliki karakteristik berbeda, yaitu XGBoost, Random Forest, dan Naïve Bayes. Pemilihan algoritma ini bertujuan untuk menganalisis bagaimana perbedaan kompleksitas model mempengaruhi hasil klasifikasi pada data yang memiliki potensi hubungan deterministik. XGBoost dan Random Forest merupakan algoritma berbasis ensemble yang memiliki kemampuan tinggi dalam menangkap pola kompleks dalam data, sedangkan Naïve Bayes merupakan algoritma probabilistik yang memiliki asumsi independensi antar fitur. Perbedaan karakteristik ini digunakan untuk mengamati bagaimana masing-masing model merespon struktur data yang digunakan dalam penelitian [5].

Untuk memastikan hasil evaluasi yang lebih reliabel, penelitian ini tidak menggunakan metode *train-test split* tunggal. Sebagai gantinya, digunakan metode *k-fold cross-validation*, di mana *dataset* dibagi menjadi beberapa subset, dan proses pelatihan serta pengujian dilakukan secara bergantian pada setiap subset. Pendekatan ini bertujuan untuk mengurangi *bias* evaluasi dan meningkatkan stabilitas hasil model [13]. Evaluasi model dilakukan menggunakan beberapa metrik performa, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Penggunaan beberapa metrik ini bertujuan untuk memberikan gambaran yang lebih komprehensif terhadap performa model, serta menghindari *bias* yang mungkin muncul jika hanya menggunakan satu metrik evaluasi. Namun demikian, dalam penelitian ini metrik performa tidak digunakan sebagai satu-satunya indikator keberhasilan model. Hasil evaluasi selanjutnya dianalisis dalam konteks validitas model, dengan mempertimbangkan potensi *data leakage*, hubungan deterministik antara fitur dan label, serta perbedaan perilaku antar algoritma dalam mempelajari data [18]. Dengan demikian, proses pengembangan dan evaluasi model dalam penelitian ini tidak hanya berfokus pada pencapaian performa tinggi, tetapi juga pada pemahaman terhadap keterbatasan model dalam merepresentasikan permasalahan klasifikasi stunting secara realistis.

2.6 Validity Analysis

Setelah Analisis validitas dilakukan untuk mengevaluasi apakah performa model *machine learning* yang dihasilkan mencerminkan kemampuan prediksi yang sesungguhnya atau dipengaruhi oleh karakteristik *dataset* yang digunakan. Tahap ini menjadi penting dalam penelitian ini, mengingat penggunaan data antropometri memiliki potensi hubungan langsung antara fitur dan label. Analisis validitas dalam penelitian ini dilakukan melalui tiga pendekatan utama, yaitu *data leakage* analysis, *deterministic relationship* analysis, dan *model behavior analysis*. *Data leakage* analysis dilakukan untuk mengidentifikasi kemungkinan adanya informasi pada fitur yang berkaitan langsung dengan proses pembentukan label. Kondisi ini dapat menyebabkan model menghasilkan performa tinggi secara semu karena memanfaatkan informasi yang sebenarnya tidak independen [18].

Selanjutnya, *deterministic relationship* analysis digunakan untuk mengkaji hubungan antara fitur dan label. Dalam konteks *dataset* ini, label status gizi ditentukan berdasarkan variabel umur dan tinggi badan sesuai standar WHO, sehingga terdapat kemungkinan hubungan deterministik yang dapat mempengaruhi proses pembelajaran model. Selain itu, *model behavior analysis* dilakukan untuk membandingkan respons berbagai algoritma *machine learning* terhadap struktur data yang sama. Perbedaan karakteristik algoritma digunakan untuk mengidentifikasi bagaimana masing-masing model mempelajari pola dalam data. Melalui pendekatan ini, analisis validitas tidak hanya

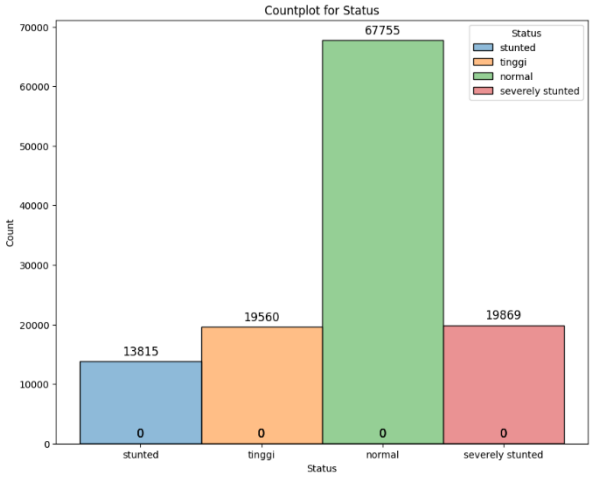
Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

berfokus pada hasil performa model, tetapi juga pada pemahaman terhadap keterbatasan model dalam merepresentasikan permasalahan klasifikasi stunting secara realistis.

3. HASIL DAN PEMBAHASAN

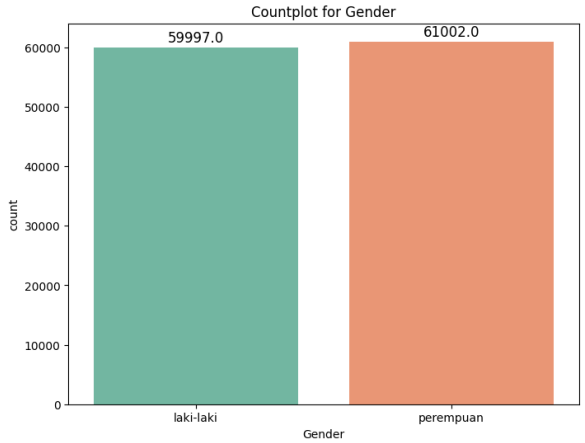
3.1 *Dataset*

Dataset yang digunakan dalam penelitian ini merupakan *dataset* antropometri balita yang berisi informasi mengenai kondisi pertumbuhan anak. *Dataset* terdiri dari 120.999 data dengan beberapa atribut utama yaitu umur (bulan), jenis kelamin, dan tinggi badan (cm). Variabel status gizi digunakan sebagai variabel target dalam proses klasifikasi. Berdasarkan hasil eksplorasi data awal, *dataset* tidak memiliki nilai yang hilang (missing value) sehingga seluruh data dapat digunakan secara langsung dalam proses analisis. Hal ini menunjukkan bahwa *dataset* telah melalui proses kurasi dengan kualitas data yang tinggi, ditunjukkan oleh tidak adanya missing value dan struktur distribusi yang konsisten untuk digunakan dalam pembangunan model *machine learning*. Distribusi data berdasarkan kategori status gizi ditunjukkan pada Gambar 2.



Gambar 2. Visualisasi Distribusi Status Gizi

Berdasarkan Gambar 2, diketahui bahwa kelas normal merupakan kelas dengan jumlah data terbanyak yaitu sebanyak 67.755 data, diikuti oleh kelas severely stunted sebanyak 19.869 data, tinggi sebanyak 19.560 data, dan stunted sebanyak 13.815 data. Distribusi ini menunjukkan bahwa *dataset* memiliki ketidakseimbangan kelas (class imbalance) yang berpotensi mempengaruhi performa model terutama pada algoritma tertentu yang sensitif terhadap distribusi data.

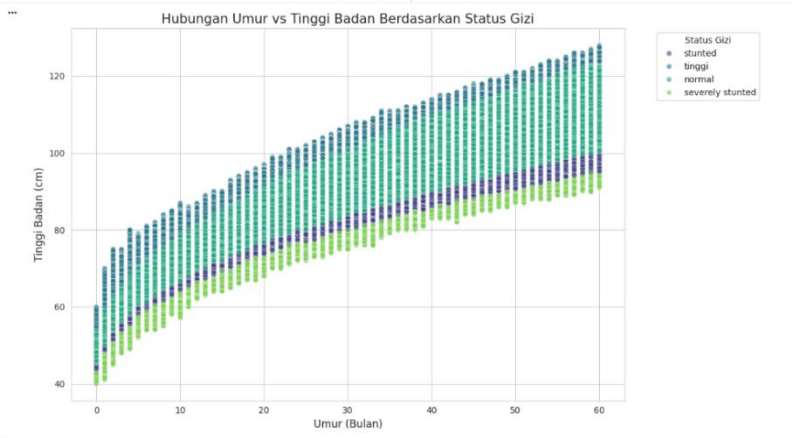


Gambar 3. Distribusi Berdasarkan Jenis Kelamin

Distribusi data berdasarkan jenis kelamin ditunjukkan pada Gambar 3. Berdasarkan visualisasi tersebut, jumlah balita perempuan sebanyak 61.002 data, sedangkan balita laki-laki sebanyak 59.997 data. Perbedaan jumlah ini relatif kecil sehingga distribusi gender pada *dataset* dapat dianggap cukup seimbang. Namun demikian, terdapat beberapa karakteristik *dataset* yang perlu diperhatikan dalam konteks validitas model. Untuk memahami hubungan antar variabel, dilakukan visualisasi antara umur dan tinggi badan berdasarkan kategori status gizi yang ditunjukkan pada Gambar 4. Berdasarkan visualisasi tersebut, terlihat bahwa setiap kategori status gizi membentuk pola distribusi yang

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

terstruktur dan cenderung terpisah satu sama lain. Kelas severely stunted berada pada rentang tinggi badan yang lebih rendah, diikuti oleh kelas stunted, normal, dan tinggi yang menempati rentang nilai yang lebih tinggi.



Gambar 4. Hubungan Umur dan Tinggi Badan Berdasarkan Status Gizi

Pola distribusi ini menunjukkan bahwa terdapat hubungan yang kuat antara variabel umur dan tinggi badan terhadap status gizi. Secara khusus, batas antar kelas terlihat cukup jelas dan minim overlap, yang mengindikasikan bahwa label status gizi kemungkinan besar ditentukan secara langsung berdasarkan kombinasi kedua variabel tersebut. Namun demikian, karakteristik *dataset* seperti tidak adanya missing value serta pola distribusi yang sangat terstruktur juga mengindikasikan bahwa data kemungkinan telah melalui proses kurasi atau mengikuti aturan tertentu. Kondisi ini berbeda dengan data kesehatan di dunia nyata yang umumnya mengandung noise, kesalahan pengukuran, serta variasi biologis yang lebih kompleks. Oleh karena itu, karakteristik *dataset* ini perlu diperhatikan lebih lanjut dalam analisis hasil model, khususnya dalam konteks validitas dan kemampuan generalisasi model terhadap data nyata.

3.2 Preprocessing Data

Tahap preprocessing data dilakukan untuk memastikan bahwa *dataset* berada dalam kondisi yang siap untuk dianalisis, dengan struktur data yang konsisten dan tanpa nilai yang hilang pada proses pemodelan *machine learning*. *Dataset* dalam penelitian ini terdiri dari empat atribut utama, yaitu umur (bulan), jenis kelamin, tinggi badan (cm), dan status gizi sebagai variabel target. Deskripsi atribut *dataset* disajikan pada Tabel 1.umur (bulan), jenis kelamin, tinggi badan (cm), dan status gizi.

Tabel 1. Deskripsi Atribut *Dataset*

No	Nama Atribut	Tipe Data	Keterangan
1	Umur (bulan)	Numerik	Menunjukan usia balita dalam bulan
2	Jenis Kelamin	Kategorikal	Jenis kelamin balita (Laki – Laki atau Perempuan)
3	Tinggi Badan (cm)	Numerik	Tinggi badan balita dalam cm
4	Status Gizi	Kategorikal	Status Gizi (Normal, *Stunted*,Suverely *Stunted*, Tinggi)

Berdasarkan hasil eksplorasi data awal, tidak ditemukan nilai yang hilang (missing value) pada seluruh atribut *dataset*. Hal ini menunjukkan bahwa data berada dalam kondisi lengkap dan tidak memerlukan proses imputasi. Hasil pengecekan missing value ditunjukkan pada Tabel 2.

Tabel 2. *Missing value*

Atribut	Jumlah *Missing value*
Umur (bulan)	0
Jenis Kelamin	0
Tinggi Badan (cm)	0
Status Gizi	0

Untuk memberikan gambaran mengenai isi *dataset*, contoh data ditunjukkan pada Tabel 3. Tabel ini merepresentasikan sebagian kecil dari keseluruhan *dataset* yang digunakan dalam penelitian.

Tabel 3. Contoh Data *Dataset*

No	Umur (bulan)	Jenis Kelamin	Tinggi Badan (cm)	Status Gizi
1	0	laki-laki	44.59	*stunted*
2	0	laki-laki	56.70	tinggi
3	0	laki-laki	46.86	normal
4	0	laki-laki	47.51	normal

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

No	Umur (bulan)	Jenis Kelamin	Tinggi Badan (cm)	Status Gizi
5	0	laki-laki	42.74	*severely stunted*
6	0	laki-laki	44.26	*stunted*
7	0	laki-laki	59.57	tinggi
8	0	laki-laki	42.70	*severely stunted*
9	0	laki-laki	45.25	*stunted*
10	0	laki-laki	57.20	tinggi

Tahap selanjutnya adalah proses transformasi data melalui teknik encoding untuk mengubah data kategorikal menjadi bentuk numerik. Variabel jenis kelamin diubah menggunakan metode *One Hot Encoding*, sedangkan variabel status gizi sebagai label target dikonversi menggunakan *Label Encoding*. Hasil transformasi data ditunjukkan pada Tabel 4.

Tabel 4. Data Setelah *Encoding*

No	Umur (bulan)	laki-laki	perempuan	Tinggi Badan (cm)	Status
1	0	1	0	44.59	2
2	0	1	0	56.70	1
3	0	1	0	46.86	0
4	0	1	0	47.51	0
5	0	1	0	42.74	3
6	0	1	0	44.26	2
7	0	1	0	59.57	1
8	0	1	0	42.70	3
9	0	1	0	45.25	2
10	0	1	0	57.20	1

Setelah proses encoding selesai dilakukan, *dataset* kemudian dipisahkan menjadi variabel fitur (X) dan variabel target (y). Variabel fitur terdiri dari atribut umur, jenis kelamin, dan tinggi badan, sedangkan variabel target merupakan status gizi balita. Namun demikian, karakteristik *dataset* yang tidak memiliki missing value serta memiliki struktur yang sangat teratur menunjukkan bahwa data kemungkinan telah melalui proses kurasi atau mengikuti aturan tertentu. Kondisi ini berbeda dengan data kesehatan di dunia nyata yang umumnya mengandung noise dan variasi yang lebih kompleks. Oleh karena itu, karakteristik ini perlu diperhatikan lebih lanjut karena berpotensi mempengaruhi hasil evaluasi model, khususnya dalam konteks validitas dan kemampuan generalisasi model.

3.3 Modeling

Pada tahap ini dilakukan pembangunan model klasifikasi untuk memprediksi status gizi balita menggunakan beberapa algoritma *machine learning*. Algoritma yang digunakan dalam penelitian ini adalah XGBoost, Random Forest, dan Naïve Bayes. Pemilihan ketiga algoritma ini didasarkan pada perbedaan karakteristik pembelajaran yang dimiliki, sehingga dapat digunakan untuk menganalisis bagaimana masing-masing model merespon struktur data yang memiliki potensi hubungan deterministik antara fitur dan label. XGBoost dan Random Forest merupakan algoritma berbasis *ensemble learning* yang mampu menangkap pola kompleks dalam data, khususnya hubungan non-linear antara variabel umur dan tinggi badan terhadap status gizi, sedangkan Naïve Bayes merupakan algoritma probabilistik dengan asumsi independensi antar fitur. Dalam penelitian ini, proses evaluasi model dilakukan menggunakan metode *k-fold cross-validation* untuk memastikan hasil yang lebih stabil dan mengurangi *bias* evaluasi. Seluruh *dataset* digunakan secara bergantian sebagai data pelatihan dan data pengujian, sehingga performa model yang dihasilkan lebih representatif.

3.3.1 *XGBoost*

Algoritma XGBoost digunakan sebagai salah satu model utama karena kemampuannya dalam memodelkan hubungan non-linear dan menangkap pola kompleks dalam data. Pada penelitian ini, XGBoost dilatih menggunakan fitur umur, jenis kelamin, dan tinggi badan sebagai variabel input. Selain itu, dilakukan proses *hyperparameter tuning* untuk memperoleh konfigurasi model yang optimal. Penggunaan XGBoost dalam penelitian ini bertujuan untuk mengamati bagaimana model dengan kompleksitas tinggi merespon struktur data yang memiliki pola yang terorganisir, serta bagaimana hal tersebut mempengaruhi performa yang dihasilkan.

3.3.2 *Random Forest*

Algoritma Random Forest digunakan sebagai model pembanding berbasis *ensemble learning* yang memiliki kemampuan dalam menangani variasi data dan mengurangi risiko *overfitting*. Model ini dibangun dengan menggabungkan sejumlah decision tree yang dilatih menggunakan subset data dan fitur yang berbeda. Dalam penelitian ini, Random Forest digunakan untuk mengamati bagaimana model dengan pendekatan agregasi mampu menangkap pola dalam *dataset*, khususnya dalam kondisi di mana terdapat hubungan kuat antara fitur dan label.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

3.3.3 *Naïve Bayes*

Algoritma Naïve Bayes digunakan sebagai model pembanding dengan pendekatan probabilistik yang mengasumsikan independensi antar fitur. Model ini dipilih untuk mengevaluasi bagaimana algoritma dengan asumsi sederhana merespon *dataset* yang memiliki keterkaitan antar variabel. Dalam konteks penelitian ini, Naïve Bayes berperan sebagai pembanding terhadap model berbasis ensemble, sehingga dapat dianalisis perbedaan performa yang dihasilkan berdasarkan karakteristik algoritma yang digunakan.

3.4 Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur performa algoritma *machine learning* dalam mengklasifikasikan status gizi balita berdasarkan data antropometri. Evaluasi dilakukan menggunakan metode *k-fold cross-validation* untuk memastikan hasil yang lebih stabil dan mengurangi *bias* yang dapat terjadi akibat pembagian data tunggal. Pada penelitian ini digunakan beberapa metrik evaluasi klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Penggunaan beberapa metrik ini bertujuan untuk memberikan gambaran yang lebih komprehensif terhadap performa model dalam mengklasifikasikan data. Hasil evaluasi model untuk masing-masing algoritma disajikan pada Tabel 5.

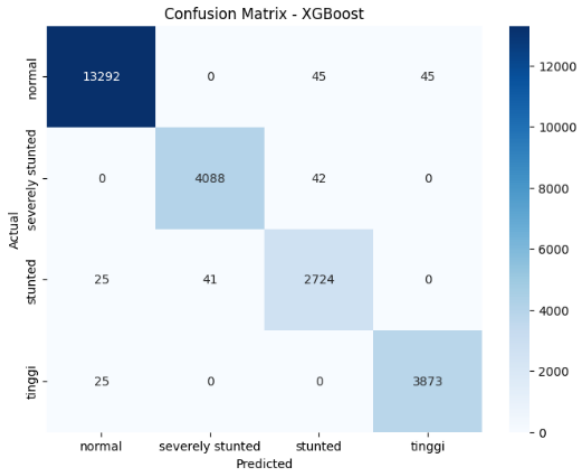
Tabel 5. Perbandingan Evaluasi Model

Model	*Accuracy*	*Precision* (avg)	*Recall* (avg)	*F1-score* (avg)
Random Forest	99.91%	1.00	1.00	1.00
XGBoost	99.08%	0.99	0.99	0.99
Naïve Bayes	55%	0.39	0.55	0.45

Berdasarkan Tabel 5, algoritma Random Forest menunjukkan performa terbaik dengan nilai *accuracy* sebesar 99,91%, diikuti oleh XGBoost dengan nilai *accuracy* sebesar 99,08%. Kedua algoritma berbasis *ensemble learning* tersebut juga menunjukkan nilai *precision*, *recall*, dan *F1-score* yang sangat tinggi dan mendekati sempurna. Sebaliknya, algoritma Naïve Bayes menunjukkan performa yang jauh lebih rendah dibandingkan kedua algoritma lainnya, dengan nilai *accuracy* sebesar 55%. Nilai *precision*, *recall*, dan *F1-score* yang relatif rendah mengindikasikan model mengalami kesulitan dalam mengklasifikasikan data secara akurat karena asumsi independensi antar fitur tidak terpenuhi.

3.4.1 Evaluasi Model *XGBoost*

Untuk memberikan gambaran lebih detail mengenai hasil klasifikasi, *confusion matrix* dari model *XGBoost* ditunjukkan pada Gambar 4.



Gambar 4. *Confusion matrix XGBoost*

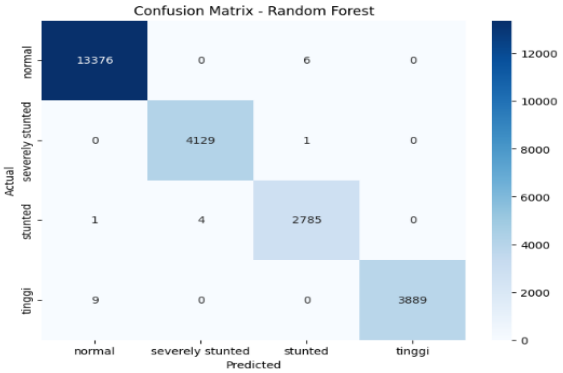
Berdasarkan *confusion matrix* pada Gambar 4, terlihat bahwa sebagian besar prediksi berada pada diagonal utama, yang menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat akurasi yang tinggi, yang mengindikasikan adanya pola yang sangat jelas dalam *dataset*. Pada kelas normal, model berhasil mengklasifikasikan sebanyak 13.292 data dengan benar, dengan jumlah kesalahan yang relatif kecil ke kelas lain. Pada kelas severely stunted, model menunjukkan hasil yang konsisten dengan 4.088 data yang terklasifikasi dengan benar. Untuk kelas stunted, model mampu mengklasifikasikan sebanyak 2.724 data dengan benar, meskipun masih terdapat beberapa kesalahan prediksi ke kelas lain. Sementara itu, pada kelas tinggi, model menghasilkan 3.873 prediksi yang sesuai dengan label sebenarnya. Secara umum, distribusi nilai pada *confusion matrix* menunjukkan bahwa model XGBoost mampu menangkap pola yang sangat terstruktur dalam *dataset*, yang mengindikasikan adanya hubungan deterministik antara fitur dan label. Namun demikian, dominasi nilai pada diagonal utama serta rendahnya kesalahan klasifikasi

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

juga mengindikasikan adanya pola data yang sangat terstruktur, sehingga hasil performa yang tinggi perlu dianalisis lebih lanjut dalam konteks validitas model.

3.4.2 Evaluasi Model *Random Forest*

Confusion matrix dari model *Random Forest* ditunjukkan pada Gambar 5.

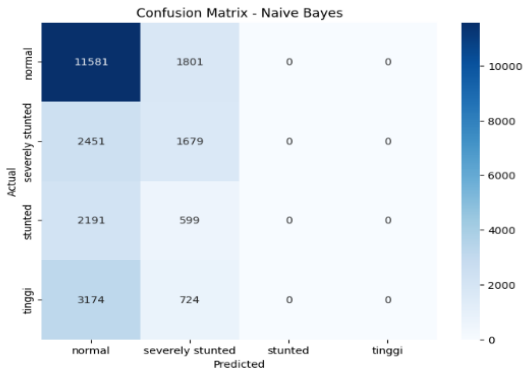


Gambar 5. *Confusion matrix Random Forest*

Berdasarkan *confusion matrix* pada Gambar 5, terlihat bahwa sebagian besar prediksi berada pada diagonal utama, yang menunjukkan bahwa model mampu mengklasifikasikan data dengan tingkat akurasi yang sangat tinggi pada masing-masing kelas, yang mengindikasikan bahwa pola antar kelas dalam *dataset* sangat jelas dan mudah dipisahkan. Pada kelas normal, model berhasil mengklasifikasikan sebanyak 13.376 data dengan benar, dengan jumlah kesalahan yang sangat kecil ke kelas lain. Pada kelas severely stunted, model menunjukkan performa yang sangat tinggi pada seluruh metrik evaluasi, yang perlu ditinjau lebih lanjut dalam konteks validitas model dengan 4.129 data yang terklasifikasi dengan benar, serta kesalahan prediksi yang hampir tidak signifikan. Untuk kelas stunted, model mampu mengklasifikasikan sebanyak 2.785 data dengan benar dengan hanya sedikit kesalahan ke kelas lain. Sementara itu, pada kelas tinggi, model menghasilkan 3.889 prediksi yang sesuai dengan label sebenarnya. Secara keseluruhan, distribusi nilai pada *confusion matrix* menunjukkan bahwa model Random Forest memiliki tingkat kesalahan yang sangat rendah dan performa yang mendekati sempurna. Namun demikian, performa yang sangat tinggi ini, terutama dengan hampir tidak adanya kesalahan klasifikasi, mengindikasikan bahwa model kemungkinan mempelajari pola yang sangat terstruktur dalam *dataset*. Oleh karena itu, hasil ini perlu dianalisis lebih lanjut dalam konteks validitas model untuk memastikan bahwa performa yang diperoleh tidak semata-mata disebabkan oleh karakteristik data yang bersifat deterministik.

3.4.3 Evaluasi Model *Naïve Bayes*

Confusion matrix dari model *Random Forest* ditunjukkan pada Gambar 6.



Gambar 6. *Confusion matrix Naïve Bayes*

Berdasarkan *confusion matrix* pada Gambar 6, terlihat bahwa model Naïve Bayes menunjukkan performa yang jauh lebih rendah dibandingkan model sebelumnya. Hal ini terlihat dari distribusi prediksi yang tidak merata, di mana sebagian besar data hanya diklasifikasikan ke dalam dua kelas utama, yaitu normal dan severely stunted, sementara kelas stunted dan tinggi hampir tidak pernah diprediksi oleh model. Pada kelas normal, model masih mampu mengklasifikasikan sebanyak 11.581 data dengan benar, namun terdapat kesalahan prediksi yang cukup signifikan ke kelas severely stunted. Untuk kelas severely stunted, hanya 1.679 data yang berhasil diprediksi dengan benar, sementara sebagian besar data salah diklasifikasikan sebagai kelas normal. Pada kelas stunted dan tinggi model tidak mampu membedakan antar kelas secara efektif, yang ditunjukkan oleh kecenderungan prediksi yang terpusat pada kelas tertentu. Hal ini menunjukkan bahwa model mengalami kesulitan dalam membedakan antar kelas secara efektif

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

karena tidak mampu menangkap keterkaitan antar fitur yang bersifat saling bergantung. Secara keseluruhan, pola distribusi pada *confusion matrix* menunjukkan bahwa model Naïve Bayes memiliki keterbatasan dalam mempelajari pola pada *dataset* yang digunakan. Rendahnya performa model ini kemungkinan disebabkan oleh asumsi independensi antar fitur yang tidak terpenuhi, mengingat variabel umur dan tinggi badan memiliki hubungan yang saling berkaitan. Perbedaan performa yang sangat kontras dibandingkan dengan model berbasis *ensemble learning* mengindikasikan bahwa karakteristik *dataset* sangat mempengaruhi hasil pembelajaran model. Oleh karena itu, hasil ini menjadi penting untuk dianalisis lebih lanjut dalam konteks validitas model, khususnya terkait hubungan antar fitur dan label yang digunakan.

3.5 Analisis Validitas Model

Berdasarkan hasil evaluasi model yang telah disajikan pada bagian sebelumnya, terlihat adanya perbedaan performa yang sangat signifikan antar algoritma yang digunakan. Model berbasis *ensemble learning* seperti Random Forest dan XGBoost menunjukkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang sangat tinggi dan mendekati sempurna. Sebaliknya, algoritma Naïve Bayes menunjukkan performa yang jauh lebih rendah pada seluruh metrik evaluasi. Jika dilihat secara lebih rinci, tingginya nilai *precision*, *recall*, dan *F1-score* pada model Random Forest dan XGBoost menunjukkan bahwa model tidak hanya mampu mengklasifikasikan data dengan benar secara umum (*accuracy*), tetapi juga memiliki kemampuan analisis dalam mengidentifikasi setiap kelas secara konsisten. Hal ini diperkuat oleh hasil *confusion matrix* pada Gambar 5 dan Gambar 6, di mana hampir seluruh prediksi berada pada diagonal utama dengan jumlah kesalahan yang sangat minimal. Namun demikian, performa yang sangat tinggi dan mendekati sempurna ini justru menjadi indikasi adanya potensi permasalahan dalam validitas model. Dalam konteks penelitian ini, variabel input yang digunakan yaitu umur dan tinggi badan merupakan variabel yang secara langsung digunakan dalam penentuan status gizi berdasarkan standar antropometri. Hal ini berarti bahwa label yang diprediksi oleh model merupakan hasil dari hubungan deterministik yang melibatkan fitur yang sama. Dengan kata lain, model tidak sepenuhnya melakukan proses pembelajaran terhadap pola baru, melainkan cenderung merekonstruksi aturan klasifikasi yang telah ditentukan sebelumnya, melainkan cenderung merekonstruksi aturan yang telah ada. Indikasi hubungan deterministik ini juga diperkuat oleh visualisasi pada Gambar 4, di mana distribusi data menunjukkan pola yang sangat terstruktur dengan batas antar kelas yang jelas dan minim overlap. Kondisi ini memungkinkan model dengan kompleksitas tinggi seperti Random Forest dan XGBoost untuk dengan mudah memisahkan kelas, sehingga menghasilkan nilai *precision* dan *recall* yang sangat tinggi pada setiap kategori. Sebaliknya, model Naïve Bayes menunjukkan nilai *precision*, *recall*, dan *F1-score* yang rendah, serta distribusi prediksi yang tidak merata pada *confusion matrix*. Model ini cenderung mengklasifikasikan data hanya ke dalam kelas tertentu dan gagal mengenali kelas lainnya. Hal ini disebabkan oleh asumsi independensi antar fitur yang tidak terpenuhi dalam *dataset* ini, di mana variabel umur dan tinggi badan memiliki hubungan yang erat dalam menggambarkan pertumbuhan anak.

Akibatnya, model tidak mampu menangkap struktur data secara efektif. Selain itu, karakteristik *dataset* yang tidak memiliki missing value serta memiliki pola distribusi yang sangat teratur juga menjadi faktor yang mempengaruhi hasil evaluasi model. Dalam praktik dunia nyata, data kesehatan umumnya mengandung noise, kesalahan pengukuran, serta variasi biologis yang lebih kompleks. Oleh karena itu, Performa model yang sangat tinggi pada penelitian ini tidak secara langsung merepresentasikan kemampuan generalisasi yang baik terhadap data dunia nyata, melainkan berpotensi mencerminkan adanya pola deterministik atau karakteristik *dataset* yang terlalu terstruktur. Dengan demikian, analisis terhadap seluruh metrik evaluasi menunjukkan bahwa performa tinggi yang diperoleh tidak dapat langsung diinterpretasikan sebagai keberhasilan model. Sebaliknya, hasil ini mengindikasikan bahwa model kemungkinan besar hanya mempelajari hubungan deterministik dalam data. Oleh karena itu, evaluasi model *machine learning* perlu dilakukan secara lebih komprehensif dengan mempertimbangkan hubungan antara fitur dan label, karakteristik *dataset*, serta perilaku masing-masing algoritma dalam mempelajari data.

4. KESIMPULAN

Berdasarkan Penelitian ini menunjukkan bahwa model *machine learning* seperti Random Forest dan XGBoost mampu menghasilkan performa yang sangat tinggi dalam klasifikasi status stunting berbasis data antropometri, dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang mendekati sempurna. Namun, hasil tersebut tidak serta-merta mencerminkan kemampuan model dalam memahami permasalahan secara substantif. Analisis lebih lanjut mengungkap bahwa performa tinggi yang diperoleh dipengaruhi oleh adanya hubungan deterministik antara fitur input (umur dan tinggi badan) dengan label status gizi, di mana label sebenarnya diturunkan langsung dari kombinasi kedua variabel tersebut. Kondisi ini menyebabkan model tidak melakukan pembelajaran pola yang bermakna, melainkan hanya merekonstruksi aturan yang telah ditentukan sebelumnya. Implikasi dari temuan ini sangat penting, khususnya ketika model dihadapkan pada data dunia nyata. Dalam praktik lapangan, data antropometri sering mengandung kesalahan pengukuran (human error), variasi biologis, serta ketidakkonsistenan pencatatan. Pada kondisi tersebut, model yang hanya bergantung pada pola deterministik berpotensi mengalami penurunan performa secara signifikan, karena tidak memiliki kemampuan adaptasi terhadap noise dan ketidakpastian data. Selain itu, penelitian ini juga menunjukkan bahwa penggunaan fitur antropometri saja tidak cukup untuk merepresentasikan kompleksitas permasalahan stunting. Faktor-faktor lain seperti kondisi sosial ekonomi, pola asupan gizi, serta lingkungan kesehatan

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 62−72
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.9584

memiliki peran yang signifikan dalam menentukan status gizi anak, namun tidak tercakup dalam *dataset* yang digunakan. Ketiadaan variabel-variabel tersebut membatasi kemampuan model dalam menghasilkan prediksi yang lebih kontekstual dan realistis. Dengan demikian, penelitian ini menegaskan bahwa evaluasi model *machine learning* tidak dapat hanya didasarkan pada metrik performa, tetapi harus mempertimbangkan validitas hubungan antara fitur dan label serta relevansi data terhadap kondisi nyata. Temuan ini memberikan kontribusi penting dalam literatur dengan menunjukkan bahwa performa tinggi dapat bersifat menyesatkan apabila tidak disertai dengan analisis validitas yang memadai.

REFERENCES

[1] UNICEF, "Survei Status Gizi Indonesia (SSGI) 2024: Kemajuan Berkelanjutan Dalam Mengatasi Malnutrisi Anak," Jakarta, Indonesia, 2025.
[2] World Health Organization, "Levels And Trends In Child Malnutrition," 2023. doi: 10.4060/cc2952en.
[3] Mundirin, Idawati, and I. Latif, "Klasifikasi Status Gizi Balita Berbasis Data Antropometri Menggunakan Random Forest," *Journal of Computer Science and Informatics Engineering*, vol. 04, no. 4, pp. 324–333, 2025, doi: 10.55537/cosie.v4i4.1202.
[4] M. I. Elim and E. Utami, "Performance Comparison Of Child Stunting Prediction Support Vector Machine Vs Random Forest With Grid Search Optimization," *JUTIF : Jurnal Teknik Informatika*, vol. 6, no. 5, pp. 5305–5319, 2025, doi: 10.52436/1.jutif.2025.6.5.5285.
[5] S. Bela, D. Widyawati, and I. R. Yunita, "Comparative Analysis Of Random Forest , Logistic Regression And SVM For Stunting Prediction Using Anthropometric Data," *Journal of Information Systems and Informatics*, vol. 7, no. 4, pp. 4271–4293, 2025, doi: 10.63158/journalisi.v7i4.1387.
[6] E. Prasetyo and K. Nugroho, "Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning Pada Label Multiclass Dengan Imbalance Data," *Jurnal Teknologi dan Informasi*, vol. 23, no. 1, pp. 1–10, 2024, doi: 10.62411/tc.v23i1.9779.
[7] B. Rao, M. Rashid, and G. Hasan, "Machine Learning In Predicting Child Malnutrition : A Meta-Analysis Of Demographic And Health Surveys Data," *International Journal of Environmental Research and Public Health*, pp. 1–15, 2025, doi: 10.3390/ijerph22030449.
[8] L. Wynants *et al.*, "Prediction Models For Diagnosis And Prognosis Of Covid-19 : Systematic Review And Critical Appraisal," *Journal BMJ*, pp. 1–22, 2020, doi: 10.1136/bmj.m1328.
[9] S. Kapoor and A. Narayanan, "Article Leakage And The Reproducibility Crisis In Machine- Learning-Based Science Leakage And The Reproducibility Crisis In Machine-Learning-Based Science," *Patterns*, vol. 4, no. 9, p. 100804, 2023, doi: 10.1016/j.patter.2023.100804.
[10] F. Chen, L. Wang, J. Hong, J. Jiang, and L. Zhou, "Unmasking Bias In Artificial Intelligence : A Systematic Review Of Bias Detection And Mitigation Strategies In Electronic Health Record-Based Models," *Journal of the American Medical Informatics Association*, vol. 31, no. March, pp. 1172–1182, 2024, doi: 10.1093/jamia/ocae060.
[11] A. Subbaswamy, C. Science, M. Hall, and N. C. Street, "From Development To Deployment : Dataset Shift , Causality , And Shift-Stable Models In Health AI," *Journal Biostatistics*, pp. 345–352, 2020, doi: 10.1093/biostatistics/kxz041.
[12] S. E. Davis, M. E. Matheny, S. Balu, and M. P. Sendak, "Perspective A Framework For Understanding Label Leakage In Machine Learning For Health Care," *Journal ofthe American Medical Informatics Association*, vol. 31, pp. 274–280, 2024, doi: 10.1093/jamia/ocad178.
[13] G. S. Collins *et al.*, "TRIPOD + AI Statement : Updated Guidance For Reporting Clinical Prediction Models That Use Regression Or Machine Learning Methods," *Journal BMJ*, 2024, doi: 10.1136/bmj-2023-078378.
[14] Kementrian Kesehatan Republik Indonesia, "Hasil Survei Status Gizi Indonesia (SSGI) 2022," Jakarta, Indonesia, 2023.
[15] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, "Key Challenges For Delivering Clinical Impact With Artificial Intelligence," *BMC Pediatrics*, pp. 1–9, 2019, doi: 10.1186/s12916-019-1426-2.
[16] D. Azriani, D. Agustian, Y. Zuhairini, I. N. Yulita, and M. Dhamayanti, "Prediction Models For Stunting At 2-Years-Old From Indonesian Newborn Population," *BMC Pediatrics*, 2025, doi: 10.1186/s12887-025-06096-4.
[17] A. Hendy, R. K. Ibrahim, S. Mohammed, F. Abdelaliem, and A. Zaher, "Supervised Machine Learning For Classification And Prediction Of Stunting Among Under-Five Egyptian Children," *BMC Pediatrics*, 2025, doi: 10.1186/s12887-025-06138-x.
[18] A. Hannun, "Measuring Data Leakage In Machine-Learning Models With Fisher Information," *UAI : Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, no. Uai, pp. 760–770, 2021, [Online]. Available: https://proceedings.mlr.press/v161/hannun21a.html
[19] A. M. A. Rahim, B. P. Hartato, A. Ridwan, and F. Asharudin, "Machine Learning-Based Approach For HIV / AIDS Prediction : Feature Selection And Data Balancing Strategy," *JAIC: Journal of Applied Informatics and Computing*, vol. 9, no. 2, pp. 338–347, 2025, doi: 10.30871/jaic.v9i2.9125.
[20] T. Hidayat, H. F. Kurniawan, and A. H. Nugroho, "Machine Learning Untuk Klasifikasi Gizi Balita Menggunakan Algoritma Random Forest," *InComTech: Jurnal Telekomunikasi dan Komputer*, vol. 15, no. 2, pp. 125–136, 2025, doi: 10.22441/incomtech.
[21] T. Yarkoni and J. Westfall, "Choosing Prediction Over Explanation In Psychology: Lessons From Machine Learning," *Journal HHS Public Access*, vol. 12, no. 6, pp. 1100–1122, 2019, doi: 10.1177/1745691617693393.
[22] World Bank Group, "Levels And Trends In Child Malnutrition," 2023. [Online]. Available: https://data.unicef.org/resources/jme-report-2023/